Sarah Alfrihidi

Predicting Hotel Booking Cancellation

Abstract

The goal of this project helps the company to predict whether the customer will cancel the booking or not. We have all the booking details like arrival_date_year, stays_in_week_nights, arrival_date_day_of_month etc of the customers from various countries.

Problem definition

Throughout the year, hotels get numerous bookings, and many of these bookings are canceled. Cancelled bookings result in a loss of revenue for the company; if management could foresee whether a hotel reservation will be canceled, it would be in their best interests to assign the room to someone else and win their business. This project seeks to use current and historical booking data to predict if a hotel reservation will be canceled.

Design

The data is provided by Kaggle.

I implemented three separate models Logistic Regression Model,

KNN Model , Decision Tree Model.

Data

The dataset contains 119390 rows, 32 columns. And This data set contains data for a city hotel and a resort hotel and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Algorithms

Models

- **Logistic Regression Model**
- **KNN Model**
- **Decision Tree Model**

Sarah Alfrihidi

Model Evaluation and Selection

I split the dataset into train and test from datasets size of 119k.

The official metric was the classification rate (accuracy).

Logistic Regression

```
Accuracy Score of Logistic Regression is : 0.9749141680018311
Classification Report :
              precision    recall  f1-score   support

           0       0.98      0.99      0.98     15894
           1       0.96      0.94      0.95      5951

    accuracy                           0.97     21845
   macro avg       0.97      0.97      0.97     21845
weighted avg       0.97      0.97      0.97     21845
```

KNeighborsClassifier

```
Accuracy Score of KNN is : 0.7206225680933852
Classification Report :
              precision    recall  f1-score   support

           0       0.78      0.87      0.82     15894
           1       0.48      0.33      0.39      5951

    accuracy                           0.72     21845
   macro avg       0.63      0.60      0.61     21845
weighted avg       0.70      0.72      0.70     21845
```

DECISION TREE

```
Accuracy Score of Decision Tree is : 1.0
Classification Report :
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     15894
           1       1.00      1.00      1.00      5951

    accuracy                           1.00     21845
   macro avg       1.00      1.00      1.00     21845
weighted avg       1.00      1.00      1.00     21845
```

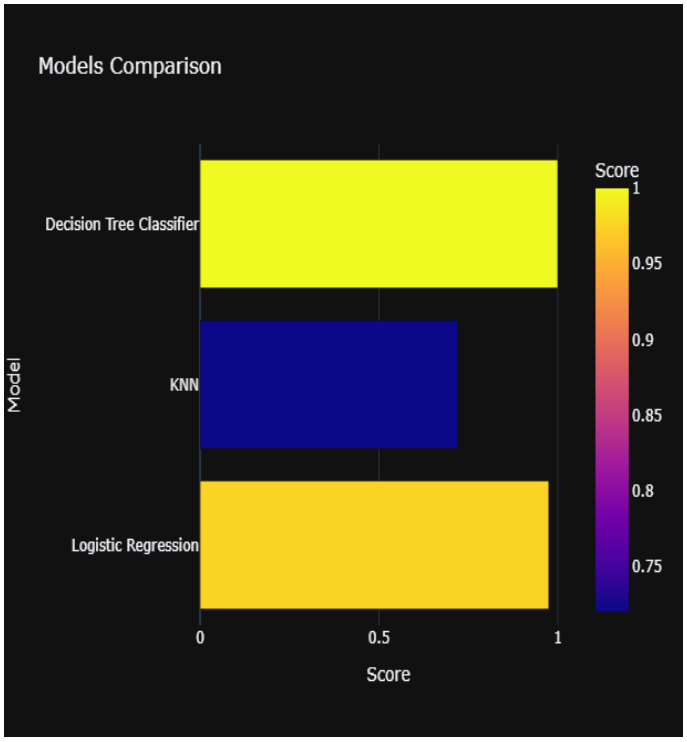
Tools

- **numpy**

- **pandas**
- **matplotlib**
- **sklearn**
- **plotly**
- **seaborn**
- **pycountry**
- **folium**

Communication

#Models comparison



Future directions

By adding more features, it will be possible to predict how long it will take for a booking to be cancelled (if at all). This would allow hotel management to develop cancellation policies and calculate cancellation fees more easily.